

Investigating Complexity in L2 Writing

Investigating Complexity in L2 Writing with Mixed Methods Approaches

Abstract

This chapter examines L2 writing from a complex systems perspective, viewing text construction as a process operating over multiple, interconnected timescales and levels. Any attempt to capture the full complexity of the system needs to be able to identify the different components in play, the timescales and levels of social organization at which they operate, the relationships between the components, and how the components and their relationships change over time. We argue, similarly to Dörnyei (2009: 109), that mixed methods research “suits the multi-level analysis of complex issues, because it allows investigators to obtain data about both the individual and the broader societal context” and discuss some of the affordances and challenges of this approach with respect to L2 writing.

“Writing occurs in writers and readers living in complex worlds.” (Bazerman, 1988: 9)

1. Introduction

L2 writing can usefully be viewed as a complex dynamic system, operating over multiple, interconnected timescales and levels of social organization. Any attempt to capture the full complexity of the system needs to be able to identify the different components in play at any particular moment and their relationships with each other. In this chapter we argue, similarly to Dörnyei (2009: 109), that a mixed methods approach, using both qualitative and quantitative measures, “suits the multi-level analysis of complex issues, because it allows

Investigating Complexity in L2 Writing

investigators to obtain data about both the individual and the broader societal context". We discuss some of the affordances and challenges of this approach with respect to L2 writing, giving examples from an ongoing project investigating L2 writing processes in Japanese learners using digital screen capture and eye tracking technologies.

2. A complex systems perspective on L2 writing

Larsen-Freeman and Cameron (2008) suggest that the modelling of any complex dynamic system should begin by identifying all of the different components in play, the timescales and levels of social organization at which they operate, the relationships between the components, and how the components and their relationships change over time. From this perspective, the dynamics of L2 writing can be represented as shown in Figure 1.

INSERT FIGURE 1 NEAR HERE

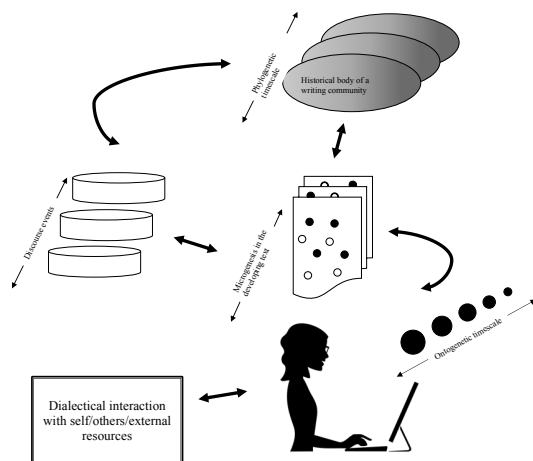


Figure 1. A complex systems perspective on L2 writing (adapted from Larsen-Freeman & Cameron, 2008)

Investigating Complexity in L2 Writing

Mirroring the principles guiding mediated discourse analysis (e.g., Wong Scollon & de Saint George, 2012), *mediated action at the site of engagement* (when the social actors are acting) is the focus of our attention. The action, in this case a L2 writing event, is considered to be ‘mediated’ because, as Scollon (2001: 4) explains:

[actions are] carried out through material objects in the world [...] in dialectical interaction with structures of the habitus. [These mediational means are taken to] be multiple in any single action, to carry with them historical affordances and constraints, and to be inherently polyvocal, intertextual, and interdiscursive [...] organized in a variety of ways, either in hierarchical structures of activities or in relatively expectable relations of salience or importance.

As an example, the writer in Figure 1 is depicted during the composition process, with the text (shown centrally as a series of ‘pages’) seen to be developing microgeneticallyⁱ, with the active language or other semiotic resources ‘in play’ at any instant (including text, images, color, font style/size (Kress, 2010)) shown on the pages of the developing text as shaded dots. The active language triggers new thoughts or preconditions later language choices (e.g., lexico-grammatical options are narrowed down as the initial part of a sentence is expressed) and these, shown as unshaded dots on the pages, lie in the background, available for activation if required. While writing, the writer is in dialectical interaction with the self, others (e.g., peers or tutors), or external resources (e.g., books, websites, or dictionaries) and these activities impact the developing discourse contingently—we never know what a final composition will look like until it is completed, and its shape will depend on exactly which components in the writer’s habitus become the focus of attention and when.

Investigating Complexity in L2 Writing

This kind of microlevel analysis of developmental processes, which CDST encourages, is relatively rare in the literature (but see, for example, Ganém-Gutiérrez, 2008; Fogal, 2019).

The mediated action of the writing event takes place in a ‘nexus of practice’ (Scollon, 2001: 4) that connects it to different timescales and levels of social organization. Writers have their own ontogeneticⁱⁱ history (represented in Figure 1 as receding shaded circles), which both shapes and is shaped by the ongoing writing process. The completed text can be seen as a distinct ‘discourse event’ which then forms part of a series of interconnected events (shown as cylinders in Figure 1), built up over a period of weeks, months, or years. For example, L2 writers might first learn how to construct paragraphs with topic sentences and supporting information, building on this to produce essays with an introduction, main body, and conclusion before moving on to look at different genres of academic essays. At a higher level of social organization, both the individuals and the discourse events they participate in form part of the ‘historical body’—“the storehouse of discourse sedimented in the history and memory of the individual and manifested in ‘habitual’ practices” (Scollon & Scollon, 2004, cited in Jones, 2008: 245)—associated with a particular community. These groups themselves change and evolve over longer, phylogeneticⁱⁱⁱ time scales, as depicted by the series of elliptical circles at the top of Figure 1. Bazerman (1988), for example, plots the history of the experimental report in the *Philosophic Transactions of the Royal Society of London* from its founding in 1665 until 1800. He shows how experimental reports began as “simply cookbook recipes for creating marvelous effects” (ibid: 66) or fortuitous observations of natural phenomena, without any obvious attempt to test hypotheses. By the end of this period, however, various recognizable features of the genre, such as identifying a research gap, making claims or counter-claims—supported by careful descriptions of experimental procedures—had become firmly established.

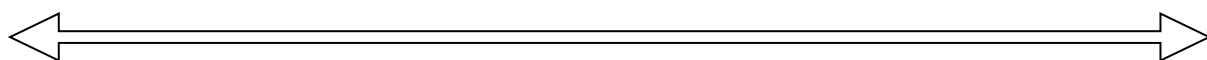
Writing can thus be seen as operating across multiple contexts of a social system, over extending time frames: from microgenesis at the micro-level, through interrelated discourse events at the meso-level, to phylogenetic development of a writing community at the macro-level. Any attempts to capture the full complexity of L2 writing will therefore need to adopt an interdisciplinary approach in order to investigate the labyrinthine networks of practice which exist and it is to this topic that we turn in the next section.

3. Mixed methods research and the study of complex dynamic systems

3.1 A brief history of mixed methods research

Mixed methods research is defined by Johnson, Onwuegbuzie, and Turner (2007: 23) as “the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches [...] for the purposes of breadth and depth of understanding and corroboration” and, in this sense, is useful for investigating a complex issue such as L2 writing. Figure 2 summarizes some of the typical characteristics of the two research traditions.

INSERT FIGURE 2 NEAR HERE



Qualitative Research

1. Naturalistic (non-interventionist)
2. Theory forming
3. Large amounts of data collected

Quantitative Research

1. Experimental (interventionist)
2. Theory testing
3. Limited amounts of data collected
4. Ideas pre-formed (‘closed technique’)

Investigating Complexity in L2 Writing

- | | |
|--|--|
| 4. No pre-conceived ideas ('open technique') | 5. Usually, all data collected is included in final account |
| 5. Only small amounts of total data included in final account | 6. High internal reliability |
| 6. Difficult to analyse independently (low internal reliability) | 7. High external reliability |
| 7. Difficult to replicate (low external reliability) | 8. Claims to be objective |
| 8. Subjective evidence | 9. Contexts less like those generalized to (lower external validity) |
| 9. Contexts resemble those the researcher wants to generalise to (higher external validity) | 10. Variables carefully controlled (higher internal validity) |
| 10. Intervening variables mean causal relationships cannot be ascribed (lower internal validity) | 11. Blind to social context |
| 11. Focuses on the social context of learning | |

Figure 2. Typical characteristics of qualitative & quantitative approaches to research (Gilmore 2007: 107)

The idea of combining quantitative and qualitative (QUAN-QUAL) methods first surfaced after WW2 and ushered in the era of the so-called 'paradigm wars', where the superiority of 'deep, rich' qualitative data was argued for by qualitative researchers while 'hard, generalizable' data was preferred by quantitative researchers (Sieber, 1973: 1335). By the late 1970s, however, the value of combining and triangulating results from both approaches in order to reach a deeper, more nuanced understanding of complex issues began

Investigating Complexity in L2 Writing

to be appreciated. A pivotal study often cited in this regard was produced by Maurice Trend (1978), a senior analyst for a research firm charged with evaluating the effectiveness of a federal housing subsidy program in the United States. His realization of the benefits of a QUAN-QUAL approach emerged largely serendipitously as a result of efforts to reconcile nonconvergent findings in the data his group collected. The difficulties in the study centered around 'Site B', where qualitative reports from the in-field observer, who judged the housing program to have been a failure, directly contradicted the quantitative measures, which suggested that Site B had performed well. Trend and a colleague spent an additional five months on a meta-analysis, redrafting their report five times, in an effort to reconcile the apparent contradictions in the findings, before finally arriving at what they regarded as a more nuanced, in-depth, and accurate explanation which satisfied the available data. In a detailed evaluation of Trend's 1978 study, Teddlie and Tashakkori (2009: 13) noted that the use of mixed methods "first allowed the opportunity for divergent views to be voiced and then served as the catalyst for a more balanced evaluation".

3.2 Mixed methods and their relationship to complex dynamic systems

Mixed methods research is now firmly established in the research community, with its own dedicated journal (*Journal of Mixed Methods Research*), and a variety of justifications have been put forward for employing its multi-strategy approach. Greene, Caracelli and Graham (1989), in a meta-analysis of 57 empirical studies, identified five possible purposes for mixed methods designs (see also Bryman, 2006 for a finer-grained analysis):

1. *Triangulation*: aims to increase the validity of results by seeking convergence, corroboration, or correspondence between different methods.

Investigating Complexity in L2 Writing

2. *Complementarity*: aims to increase the interpretability, meaningfulness, and validity of results by exploiting one method to elaborate, enhance, illustrate, or clarify another.
3. *Development*: aims to increase validity by using the results from one method to develop or inform another.
4. *Initiation*: aims to increase the breadth or depth of results by analyzing them from different perspectives and identifying paradoxes or contradictions.
5. *Expansion*: aims to increase the scope of inquiry through the use of multiple methods.

A complex systems perspective tends to increase the scope of an enquiry and is therefore close in its aims to 'initiation' or 'expansion' in Greene et al.'s (1989) list. However, the large number of interacting components in play in complex systems, operating over varying timescales and levels of a social system, necessitate the deployment of multiple methods in order to capture a fuller picture of development. In exploring L2 writing processes, investigations might include, for instance, examining moment-by-moment microgenetic changes with digital screen capture/eye-tracking technologies and think aloud protocols or tracking learners' development of writing competence over longer periods of time with a diachronic corpus.

The study of complex dynamic systems tends to align itself with the pragmatic school of thought that evolved out of the paradigm wars, prevalent during the 1970s and 1980s. Pragmatists contend that a false dichotomy exists between quantitative and qualitative approaches and that, rather than being bipolar, they both lie on a continuum of scientific enquiry, each with its own strengths and weaknesses, and each seeking to reach some level of understanding through its own means (Newman & Benz, 1998; Onwuegbuzie & Leech, 2005). An increasing number of researchers who adopt a complex systems perspective in

their work are also encouraging a pragmatic approach using mixed methods, as the following selection of quotations from varied sources suggest, and we strongly support this position:

- “Pragmatic researchers [...] are more able to combine empirical precision with descriptive precision (Onwuegbuzie, 2003). Also, armed with a bi-focal lens (i.e. both quantitative and qualitative data), rather than with a single lens, pragmatic researchers are able to *zoom in* to microscopic detail or to *zoom out* to indefinite scope (Willems & Raush, 1969). As such, pragmatic researchers have the opportunity to combine the macro and micro levels of a research issue.” (Onwuegbuzie & Leech, 2005: 383)
- “Analysis or investigation of discourse from a complex systems perspective does not require us to throw away other approaches and their techniques. Indeed, multiple types of analysis are needed to work with information from systems at different scales, and new ways of blending methods [...] are needed to explore simultaneous activity on several scales.” (Larsen-Freeman & Cameron, 2008: 194)
- “At the level of translating epistemological concerns into research methodology and finally the decision of research methods, a pragmatic paradigm, poses some methodological questions. If phenomena have different layers how can these layers be measured or observed? Mixed methods research offers to plug this gap by using quantitative methods to measure some aspects of the phenomenon in question and qualitative methods for others.” (Feilzer, 2010: 8)

3.3 The challenges associated with mixed methods research

Investigating Complexity in L2 Writing

Bryman (2007) notes that serious problems face mixed methods researchers attempting to integrate quantitative and qualitative data in practice, and highlights nine potential barriers, including: (1) Researchers privileging one kind of data over another either because of audience expectations or because of their own particular preferences; (2) Interdisciplinary issues brought about by bringing together specialists from different fields (for an excellent example of this, see Austin, Park & Goble, 2008); and (3) Publication difficulties caused by methodological bias from editors or length restrictions (mixed methods reports are inevitably longer given their dual focus).

While justifications for a mixed methods approach to investigating complex systems appear persuasive and are widely endorsed (see Section 3.2), significant obstacles also lay in the path of complexity science researchers at different stages of the research process. The first challenge involves transcending our own cognitive styles and methodological preferences by (1) gaining new skills or knowledge that might take us out of our comfort zones, or (2) forming interdisciplinary collaborative groups with the necessary specialist knowledge in order to apply a range of QUAN-QUAL research tools to any particular area of investigation. The second challenge lies in the implementation of a project: finding the necessary time and resources to enable a detailed description of the components at play in a complex system and their relationships with each other. The third challenge is at the stage of analysis: triangulating the extensive data generated with a mixed methods approach and reconciling any contradictions raised by the multiple perspectives embraced in the experimental design. The final challenge echoes Bryman's (2007) concerns about publishing, and is at the dissemination stage: identifying suitable publication outlets for interdisciplinary work, and finding ways to slice up data into meaningful 'packages' that respect the restrictive word limits typically imposed by publishers while still illustrating the richness of the complete data set.

The remainder of this chapter will explore some of the issues raised above in more detail by showing how a mixed methods approach to studying the complexity of L2 writing can work in practice. It is intended to be a ‘warts-and-all’ description of an ongoing project, describing both the challenges faced along the way and the affordances the approach has provided. The hope is that this will act as an exemplar and as a jumping-off point for other researchers interested in pursuing a similar approach. Because of space limitations, it is not possible to provide a comprehensive account (which can be found in the articles cited); rather, aspects of the project which illustrate some key features relevant to a complex systems, mixed methods perspective on L2 writing will be highlighted.

4. Applying a mixed methods approach to the investigation of L2 writing processes

The research described here, carried out at two universities in Japan, sought to provide a detailed description of the L2 writing processes of 22 students at different levels of English proficiency (see Gánem-Gutiérrez & Gilmore, 2018a; Gánem-Gutiérrez & Gilmore, 2018b; Gilmore & Gánem-Gutiérrez, forthcoming). In the project, following Larsen-Freeman and Cameron (2008), we attempted to model as much of the complexity of the emergent L2 writing process as possible with innovative uses of research tools (digital screen capture, key-stroke logging, eye tracking technology, and retrospective think-aloud protocols), along with a wide variety of quantitative and qualitative measures (learner proficiency, essay length and quality, allocation of time to different writing processes, on-screen eye movements and gaze duration, non-verbal behavior during the writing task, use of online lexicographic tools, and student retrospective reflections on their writing).

Investigating Complexity in L2 Writing

The main data for the project was collected using an eye tracker (Tobii T60/Studio 2.2) which, while allowing us to capture patterns in the L2 writing process at macro-, meso- and micro-levels, meant that each participant had to be assessed one at a time, individually. Data gathering took place in three phases: a pre-composition phase, a composition phase, and a stimulated retrospective recall (SRR) phase. At the pre-composition phase, students completed a 116-item C-Test (a variety of cloze test) in order to estimate their English proficiency (Gilmore, 2011). At the composition phase, participants were first familiarized with the hardware and software available to them during the writing task (including MS Word and various online lexicographic resources) and the eye tracker was calibrated. They were then given 45 minutes to answer an International English Language Testing System (IELTS)-style argumentative essay on the topic: “Education should be free for everyone. To what extent do you agree or disagree with this statement?” Digital screen capture (DSC) data with eye-gaze overlaid and video recordings of the participants’ interaction with the computer (13 hours of each in total) were collected which provided detailed, moment-by-moment information on the composition process. All completed essays were blind-rated by three English L1 teachers across four dimensions (task achievement, coherence and cohesion, lexical resources, and grammatical range/accuracy) using IELTS writing band descriptors. These were then averaged to produce a global score of essay quality for each composition. In the final phase, SRR data was collected after a 10-minute break (while the writing event was still fresh in the participants’ memories), with the recorded activity from phase 2 providing the stimulus (approximately 27 hours of recorded data was collected across all 22 students). Using ELAN annotation software (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006), students’ allocation of time to different writing processes (text construction, revising, rereading, use of external resources, and pausing) was calculated and their reflections on the

Investigating Complexity in L2 Writing

writing process in the SRRs were transcribed in full. Figure 3 provides an overview of all the visual data available for analysis.

INSERT FIGURE 3 NEAR HERE

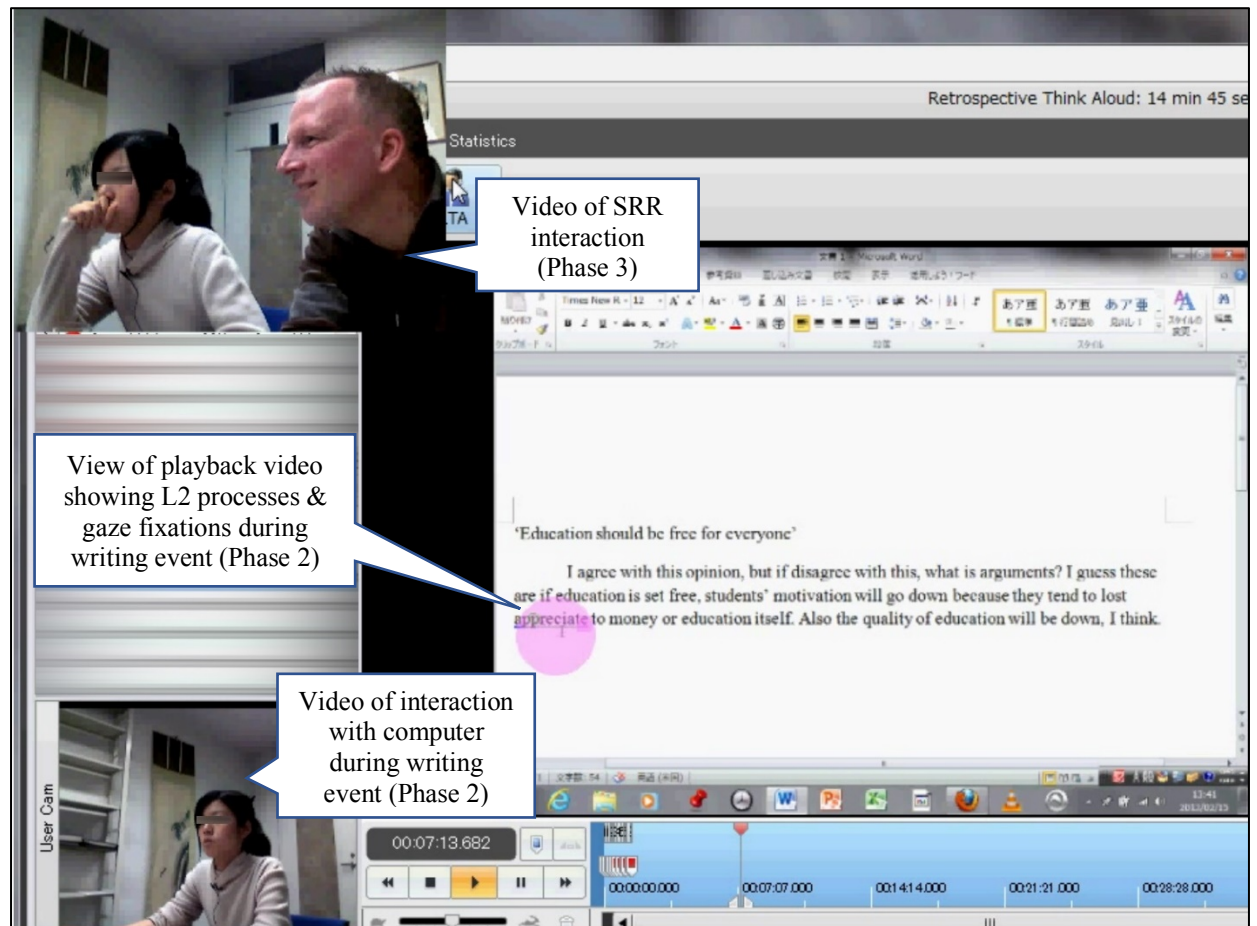


Figure 3. Overview of visual data available for analysis (Gánem-Gutiérrez & Gilmore, 2018b: 480)

In the remainder of this section, we will discuss the affordances and challenges encountered in the project as a result of our decision to employ a mixed methods approach to investigating L2 writing processes.

4.1 Methodological affordances: Investigating different levels of the complex system

Investigating Complexity in L2 Writing

The careful collection of a wide range of quantitative and qualitative data in the investigation allowed us to examine different levels of the L2 writing system. Naturally, there could always be more data to collect: (i) Since the participants self-selected into the study we had little information on their ontogenetic histories or L2 writing developmental pathways and no follow-up was possible; (ii) Only one task (argumentative essay) was assessed although, since there is evidence of task difficulty affecting strategic use of writing processes (e.g., Miller, Lindgren & Sullivan 2008), multiple task-types are needed to provide a clearer picture of writing changes; (iii) Logistical reasons meant that it was impractical to measure changes in the target population over longer, phylogenetic timescales (i.e., their complete 4-year university education). However, even as implemented the study imposed considerable demands on the researchers and participants (see Section 4.2) and a cut-off point regarding what is achievable will inevitably be reached in any investigation. Despite these limitations, in the proceeding section some examples of insights at macro-, meso- and micro-levels are given to illustrate the potential of a mixed methods approach to unpack L2 writing systems.

4.1.1 The macro-level

The application of inferential statistics in the study allowed us to pool data from individual participants and, for statistically significant findings, to generalize out to a wider population, thus providing a macro-perspective on the use of L2 writing processes in the target group. The results indicated that during the L2 writing task, in terms of frequency, text construction and revising were the dominant processes, followed by pausing and rereading, and that the overall pattern for their use was ‘little and often’, suggesting a complex approach to composing where the various writing processes are dynamically and contingently intertwined (e.g., Bereiter & Scardamalia 1987). Regarding the use of external resources (mono/bilingual dictionaries, thesauruses, or Google), these tended to occur less frequently but for longer

average durations (see Figure 4). Important here was our realization that broadening the research agenda to include both duration and frequency data revealed different characteristics of the system which would otherwise have remained hidden.

INSERT FIGURE 4 NEAR HERE

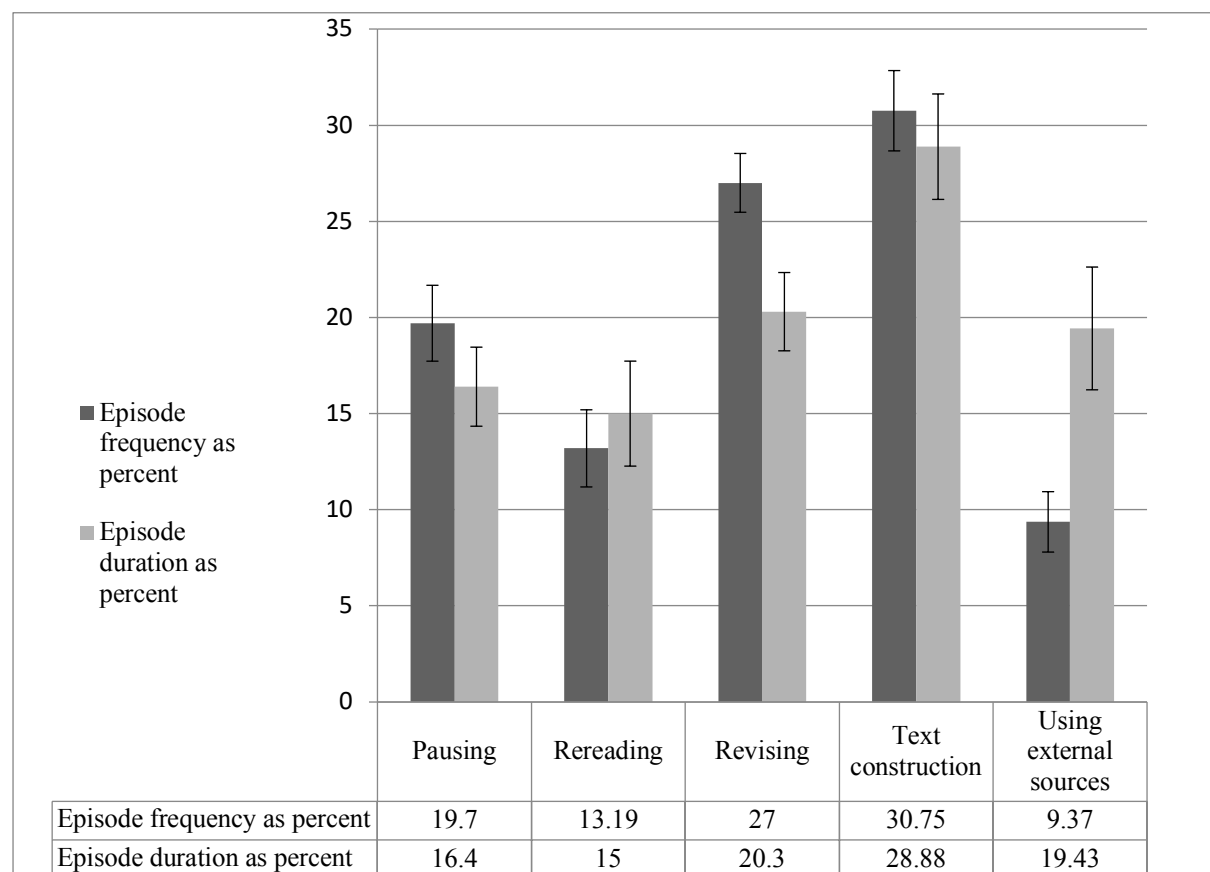


Figure 4. Frequencies and duration of episodes for each writing process type (Gánem-Gutiérrez & Gilmore, 2018b: 484)

Statistically significant interactions were found between writing process type and learner proficiency, in terms of both episode frequency ($F(4, 145) = 8.15, p < .001, \eta p^2 = .09$) and duration ($F(4, 156) = 5.72, p < .001, \eta p^2 = .10$). Follow-up Pearson correlations showed significant positive correlations between proficiency and (i) text construction, and (ii)

revision (in terms of frequency rather than duration), but a significant negative association between proficiency and use of external resources (see Table 1).

INSERT TABLE 1 NEAR HERE

	Episode frequency		Episode duration	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Text construction	.253	.013	.219	.022
Revising	.230	.016	.080	.408
Pausing	-.158	.099	-.076	.428
Rereading	.002	.979	.154	.108
Using external resources	-.342	<.001	-.321	<.01

Table 1. Correlations of proficiency with episode frequency and duration, by process type (Gánem-Gutiérrez & Gilmore, 2018b: 489)

In other words, higher proficiency students tended to write more, revise more frequently (but not for longer durations overall), and to rely less on the support of external lexicographic resources during the L2 composition task. This in itself might not seem particularly surprising, however, as Section 4.1.3 reports, the statistical results masked some interesting patterns which only became apparent at the micro-level of analysis. Thus, research methods which tap into different levels of a complex system tend to reveal different characteristics and this underscores the value of mixed methods research.

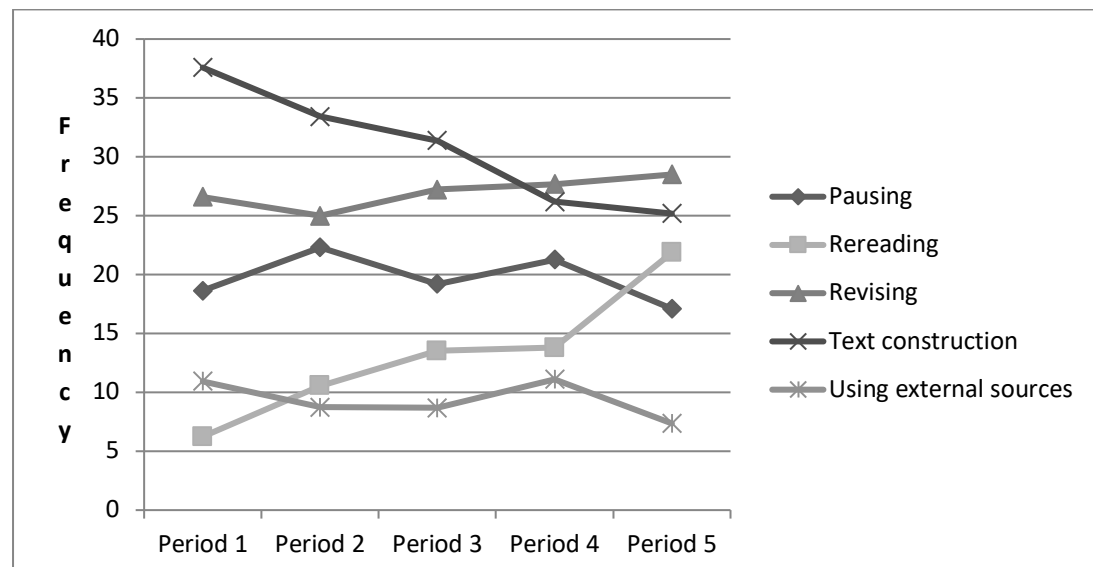
4.1.2 *The meso-level*

To investigate the meso-level of the L2 writing system in more detail, we divided the total time of the composition task into five periods in order to see whether the use of different writing processes varied in duration and/or frequency during different stages of the task. This

Investigating Complexity in L2 Writing

temporal dimension of composition activity has been rather neglected in the literature as researchers have tended to rely on indirect methods of data collection such as questionnaires, concurrent think-aloud protocols, or stimulated retrospective recall which renders this kind of data inaccessible. However, the use of digital screen capture and eye-tracking technology in our study allowed us to track the moment-by-moment deployment of different writing processes throughout the composing period. In this way we gained a better understanding of the complex cognitive activity occurring as the task unfolded. The results indicated significant variation in both the frequency and duration of the different processes across successive periods as shown in Figure 5.

INSERT FIGURE 5 NEAR HERE



Investigating Complexity in L2 Writing

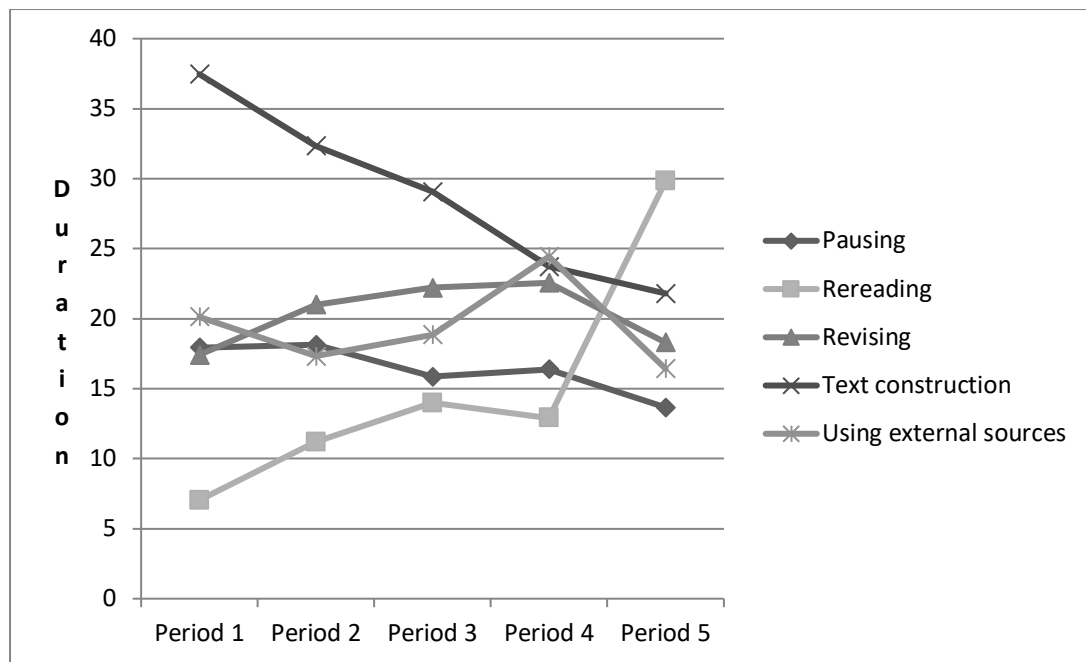


Figure 5. Frequencies and duration for each writing process type across successive periods (Gánem-Gutiérrez & Gilmore, 2018b: 486/7)

Text construction and revising were the dominant writing processes during the composition period. The clearest patterns observed were the linear trends for text construction and rereading: the former falling and the latter rising across periods 1 to 5. Revising, pausing, and use of external resources all fluctuated during the task but showed no significant tendency to rise or fall (except for a peak in the duration of external resources in Period 4). Overall, our findings supported earlier claims that different writing processes “did not stand an equal chance of being activated at any given time in the composing process” (Manchón, Roca de Larios & Murphy 2009: 108), with the task demands at any particular point affecting the choice contingently. Interestingly, once again the patterns observed in writing processes varied depending on whether frequency or duration data was the key measure, highlighting the value of a mixed methods approach in order to gain a more nuanced understanding of L2 writing development.

4.1.3 *The micro-level*

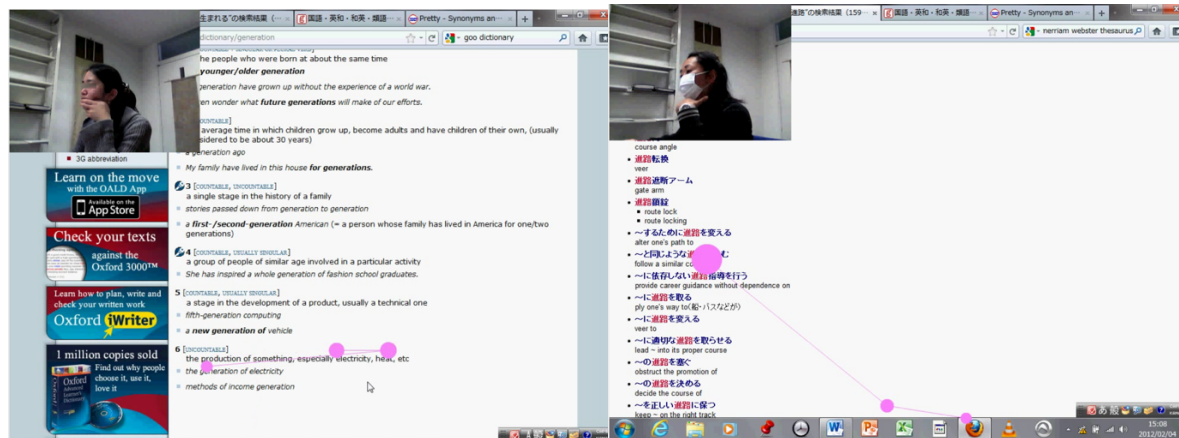
Digital screen capture, eye-tracking technology, and video recordings of writer activity during text construction or stimulated retrospective recalls (SRRs) all facilitated the detailed analysis of the L2 composition process as well as learning opportunities at the micro-level. Two examples of this analysis are illustrated in this section.

Example 1

As noted in Section 4.1.1, more proficient writers in the study tended to consult external lexicographic resources less both in terms of frequency and duration, but a micro-analysis of learner behavior during look-ups (i.e., when learners used an online resource) suggested marked qualitative differences in *how* consultation time was actually used. Specifically, the higher proficiency participants demonstrated more sophisticated and highly regulated strategic behavior and therefore tended to achieve more in a given amount of time. Figure 6 contrasts the look-up behavior for P01 (the highest proficiency learner) with P22 (the lowest proficiency learner) over similar durations of around 2 minutes, along with the two participants' reflections on their activity from the SRR sessions. Both online lexical queries by participants are triggered by a need to translate an L1 concept into the L2: 'accrual/emergence' for P01 and 'career options/paths' for P22.

INSERT FIGURE 6 NEAR HERE

Investigating Complexity in L2 Writing



High L2 proficiency (P01)			Time (secs)		Low L2 proficiency (P22)		
Action	Outcome	SRR commentary	P01	P22	Action	Outcome	SRR commentary
1. Query for <i>hassei</i> (accrual/emergence) in bilingual dictionary	Unsuccessful search	P01: it has lots of meanings but it's like the development of our feelings or something although I think I looked at a lot of definitions here	0	0	1. Query for <i>shinro</i> (career options/paths)		Tutor: what are you reading here? P22: <i>shinro</i> many many Tutor: many choices yeah
2. Skips to page 8 in the search results for <i>hassei</i>	Unsuccessful search	Tutor: skipped to page eight? P01: because I do that often and I thought that the first the second page I looked at was using <i>hassei</i> in a different sort of meaning from I was looking so I skipped to see how it's used	32	31	2. Multiple rereadings of one particular example using 'path' in the bilingual dictionary (indicated by eye gaze data and cursor movements)		P22: <i>kore dake</i> (just this one) only I think Tutor: do you think this is a good meaning? P22: mm mm mm
3. Second query, <i>umareru</i> (accrual/emergence)	Unsuccessful search	P01: I changed it again to <i>umareru</i> <i>hassei</i> is more difficult word than <i>umareru</i> so I thought <i>umareru</i> might have a better translation than <i>hassei</i>	48	56	3. Movement of cursor to Word icon at the bottom of the screen (indicated by eye gaze data)		
4. Return to Word to write 'gene' (incomplete form of generation)	Triggers new search in monolingual dictionary	P01: [triggered by DSC video showing text construction of 'gene'] oh and then I thought of the verb generate and then I changed it to generation but I thought that generation had different meanings than generate	82	57	4. Returns to examining the translations for <i>shinro</i> in the bilingual dictionary, focusing on examples using 'course'		P22: <i>kono hen wa sugoi atteru</i> (these are really appropriate) good meaning
5. Query for 'generation' in monolingual dictionary	Successful search	P01: it says that the production of something especially electricity so I thought it wouldn't fit in my [essay] So I thought I couldn't use the word generation here	88	135	5. After eye gaze fixation on the word 'course', she quickly returns to Word to insert lexical item	Inappropriate text revision	
6. Return to word processor to delete 'gene'	Successful text revision	P01: then I couldn't find the right word so I think I use a different word [laughs]	113				

Figure 6. Online strategic behavior for highest and lowest proficiency students over similar time periods

(Gánem-Gutiérrez & Gilmore, 2018b: 495/6)

P01 begins her online activity with a search in the Japanese-English bilingual dictionary Eijiro ALC for the word *hassei* (accrual/emergence) to complete a developing idea in her essay: 'Education plays a decisive role in the _____ (of people's ways of thinking)'. As she explains in the SRR, *hassei* has many possible translations which differ from her intended meaning, and she quickly scans through the first two pages of search results, looking for an appropriate translation. She then skips to page 8 of the search results in an attempt to find a translation closer to her intended meaning which ultimately proves

Investigating Complexity in L2 Writing

unsuccessful. At this stage, she alters the search word to *umareru* (a near synonym of *hassei*) in order to find a more appropriate translation and the eye tracking indicates that her attention rests briefly on an example sentence containing ‘be generated from ~’ which she seems to decide to use in her essay in the noun form (i.e. *‘Education plays a decisive role in the *generation* (of people’s ways of thinking)’). P01 returns to MS Word and types ‘gene’ before pausing, unsure (as she explains in the SRR) whether the two forms, *generate* and *generation*, have the same meaning. She then checks the meaning of the target word in an online monolingual dictionary and, discovering that it refers to ‘the production of something, especially electricity’ (see screenshot for P01 in Figure 6), she finally decides to express her idea in a different way and deletes the unfinished sentence.

In contrast, P22 spends a similar amount of time online searching ALC for an appropriate translation of a single target word, *shinro* (career options/paths). She appears overwhelmed by the number of choices offered and is slow to process the information in the search results, with the digital screen capture and eye gaze data indicating multiple re-readings of the Japanese and English translations for one example sentence, ‘Haru was worried about which path to choose, and he decided to pay a visit to his art teacher’, running the cursor arrow under the text as a reading aid. After this, the cursor movement and gaze fixation on to the Word icon at the bottom of the screen suggest that she plans to use *path* in her essay, but she then seems to have second thoughts and instead continues her examination of the search results. Scrolling down the page further, she concentrates her attention on example sentences containing the word *course*, which she comments in the SRR seemed to match her intended meaning. Finally, she fixates on this target word one last time (see screenshot for P22 in Figure 6) before quickly returning to her essay to continue typing, presumably because of short-term memory limitations. This is a pattern seen frequently in the lower proficiency students where the cognitive demands of the spelling task are high and they

appear to be unaware of the copy/paste short-cuts available for MS Word. Despite the length of time spent searching the online dictionary, P22's final text, *‘After that, they choose the course for a dream’, is relatively unintelligible to readers.

Example 2

In a separate analysis at the micro-level (Gánem-Gutiérrez & Gilmore, 2018a), using data from the same study, we focused on a single participant (P17) and investigated the extent to which tutor-student interaction during the stimulated retrospective recall supported developmental opportunities. The SRR video was first segmented into episodes of verbal and non-verbal activity using ELAN annotation software and all verbal interaction was coded as either: (i) procedural (transactional), where the learner and tutor discuss the writing event but there is no pedagogical focus; or (ii) languaging (developmental) where language is used to mediate cognitive activity and opportunities for knowledge construction/enhancement emerge (e.g., Swain, 2010). Table 2 shows the results of this analysis.

INSERT TABLE 2 NEAR HERE

Episode type	Non-verbal	Verbal		Total
		<i>Procedural</i>	<i>Languaging</i>	
No. episodes % (raw figures)	53% (98)	37% (67)	10% (19)	100% (184)
Time %	36.6%	35.9%	27.5%	100%

Table 2 Tutor-student activity during SRR (Gánem-Gutiérrez & Gilmore, 2018a: 28)

Investigating Complexity in L2 Writing

As can be seen, although languaging episodes constituted only 10% of the total number of occurrences, they tended to last longer than non-verbal or procedural episodes (making up 27.5% of total duration) and we therefore concluded that there were considerable opportunities for L2 development during the SRR. Exactly *how* learning affordances can emerge during retrospective recall was explored through the detailed examination of just one of the languaging episodes (~73 seconds duration) using *microgenetic multimodal interaction analysis*. This approach focuses on the various semiotic resources mobilised to mediate communication, understanding, and development including language, nonverbal behavior (gaze, facial expression, gesture, head and body movement, and orientation), tools (e.g., computer or paper notes), settings (e.g., tutor's office), roles and relations (e.g., expert-novice), and situated activity systems (e.g., goals, practices) (Nishino & Atkinson, 2015: 41-42).

The episode is triggered when P17 writes *‘they tend to lost appreciate to money’ in her essay and *appreciate* is underlined by MS Word’s autocorrect feature (see Figure 3). The green squiggle catches her attention as she is rereading her sentence and she fixates on this part of the screen (indicated by the pink circle of the eye tracking software), before right-clicking to bring up the grammar checker. The student (S) selects the verb form *appreciating* from the choices offered (which is also incorrect in the given context) and this causes the tutor (T) to pause the video playback, thus creating space for metacognitive and metalinguistic activity to take place, as shown in the transcript below^{iv}:

- 1 T ((T and S watching video, then T stops playback) °yeah°
- 2 actually (0.2) yeah the, you want the noun here don't you.
- 3 S [um::
- 4 T [which is? ((turns to look at her and waits)) (1.2)

Investigating Complexity in L2 Writing

5 S ((S adopts pensive gesture: tilting head, hand goes to
6 chin, distant gaze)) a (3.8) ppreciate ((turns to look
7 at T as soon as she finishes uttering the word))
8 T (2.1)that's a verb (0.5)
9 S ah ((nods slightly))
10 T yeah (0.8)
11 S apprecia::(.)tion? ((S turns to T as she utters last
12 syllable and giggles))
13 T uh hum: yeah ((nodding keenly)) that's it ((T smiles and
14 they laugh together)) (0.9) so actually [the the the
15 ((pointing at the screen)) um the suggestion it gave you is
wrong yes (0.7)
16 S [um
17 ah::=
18 T =yeah um
19 S yeah ((T laughs softly)) I've (0.3) yeah ((pointing at
20 screen)) I (1.0) apprecia:ting is (1.0) I (0.9) haven't
21 (.)seen the appreciating=
22 T =um
23 S So (0.5)I (0.2)um (0.6)appreciate (1.3) I think it's (.)
24 verb((turns to T))[and or verb and (0.6) noun?=
25 T [yeah
26 =yeah (.) uh huh
27 S ((giggles)) but ((giggles)) it's ((giggles and points to
28 screen)) (1.5) computer suggest me ((laughs))=

Investigating Complexity in L2 Writing

29 T ((turns to S)) =ye::ah it it it was (.) it gave you bad
30 advice ((smiles and laughs with her))(.)yeah appreciating
31 is just the verb the -ing form of the verb yeah

An in-depth analysis of the linguistic, paralinguistic (e.g., body language, gestures, facial expressions, and intonation), affective, and material mechanisms or tools mediating L2 development displayed in this excerpt is beyond the scope of this chapter. However, Gánem-Gutiérrez and Gilmore (2018a) provide more detail and clearly illustrate the potential insights into learning available through this approach. For example, pointing gestures used by the interlocutors seem to have both communicative and self-regulatory cognitive functions in this case.

4.2 Methodological challenges

The mixed methods approach, employed in this study to investigate L2 writing as a complex system, brought with it challenges at each stage of the research process.

4.2.1 The preparation stage

To capture nuances in the L2 writing process, the methodological choices made at the planning stage meant that we had to broaden our research skill set and collaborate with experts from other areas, including statisticians, eye-tracker technicians, and conversational analysts. In terms of research tools, the Tobii eye tracking hardware/software and ELAN annotation software were both unfamiliar to us and took a considerable amount of time to master, particularly since we relied predominantly on technical manuals for guidance. We also sought specialist advice on some of the statistical procedures and multimodal interaction analyses.

4.2.2 *The implementation stage*

Implementation of the project was complicated by the fact that all the data was collected on an individual basis for each participant. The eye tracker and associated software are costly so we only employed one Tobii T60, and this had to be recalibrated for each user before the writing task began. The total time for initial data collection, including a background questionnaire on learning history, English proficiency test (C-Test), writing task, and SRR protocol, was over two hours per student. The eye tracking software also crashed during the digital screen capture for two participants and therefore failed to record their writing activity, meaning they had to be excluded from the study.

4.2.3 *The analytical stage*

The data analysis period was extremely long and complex because of the multiple strands designed into the study:

1. The DSC videos for the original writing events had to be segmented into episodes and coded by multiple raters according to writing process type (text construction, revising, rereading, use of external resources, and pausing), and inter-rater reliability needed to be calculated because of the inherent subjectivity of some of the coding decisions.
2. The stimulated retrospective recall videos had to be transcribed in full with support from a Japanese translator to confirm the meaning of any L1 utterances which arose during the interviews, particularly with lower proficiency students. Transcription work is typically time-consuming—even with native-speaker dyads it is estimated to take five times the duration of the interaction (Johnson, 1995)—and the strong accents and dysfluency of some of the participants lengthened this process. Thus, the total transcription process took around 200 hours, with a research assistant producing the

initial rough version which was then revised by a L1 Japanese speaker and finally checked for accuracy by the researchers.

3. Understanding the online behavior of students during the lexicographic look-ups involved triangulating data from four sources: the writing event videos, the SRR videos, the transcriptions of student-tutor reflections, and the final essays produced by the writers. This process produced a final qualitative description of moment-by-moment online activity which ran to approximately 30,000 words.
4. The quality of the final IELTS-style essays had to be blind-rated by three qualified language teachers using IELTS Task 2 writing band descriptors, and inter-rater reliability had to be calculated to provide a more objective estimate of ‘task success’ for each participant.
5. Microgenetic multi-modal analysis involved the careful examination of selected SRR video extracts, incorporating all the modes and media used for meaning-making and emerging learning opportunities in the tutor-student interaction. These were recorded in transcripts using procedures borrowed from conversation analysis conventions. The fine detail required in this approach inevitably limits the quantity of data that can be reported, but the ‘snapshots’ captured in this way added invaluable depth to the study.

4.2.4 The dissemination stage

The write-up for the different strands of the study was delayed because of the lengthy analytical and triangulation work preceding it—the latter a particular feature of mixed methods research. Once the data had been processed, there was far too much information for a single journal article and we therefore planned to present the results in a series of related articles which, when taken together, would build a more complete picture of how L2 writing

Investigating Complexity in L2 Writing

events unfold. However, this approach was criticized by some anonymous reviewers as ‘salami slicing’:

In the introduction [...] you mention that the results of this investigation will be reported ‘in a series of papers’... I admire your honesty in reporting this and can fully understand why you would need to write multiple articles to present all aspects of your data and answer the most important/relevant research questions about them. However, please be aware that many journals, including the APA publishing guidelines, actively discourage publishing multiple articles based on the same experimental dataset, which they call ‘salami slicing’. I have a similar dataset to yours and have, unfortunately, found it difficult to publish multiple papers in some (APA related) journals as a result. That is why I think it is important that you consider whether you should state so explicitly that you will be publishing multiple papers about this dataset. And if you do mention it, perhaps you should provide stronger/clearer arguments for why you are going to do this.

At the same time, other reviewers complained that our work failed to do justice to the full dataset and requested more detail, particularly in terms of qualitative support from the eye-tracking and SRR strands:

The value of your current article lies precisely in its use of multiple measures. And the main issue, as I see it, is that the current version of your article does not do justice to your dataset. The presentation and analyses of all relevant data, especially related to eye-tracking and retrospective verbalization, is superficial and lacking in detail—both in terms of methodology and associated analyses. If you remove these data [...] your report will suffer in quality, as its strength is precisely in the multiple measures employed by you. I’d like you to strongly consider, in fact, the opposite—bolstering

these aspects of your article, adding to the methodological detail and to the clarity of analyses.

Ultimately, for this particular submission, we were able to negotiate for a longer article, but this sample feedback illustrates some of the tensions between breadth and depth which inevitably arise when reporting mixed methods results.

5 Conclusion

The mixed methods study outlined in this chapter has provided an incredibly rich data set which we are still analyzing and which will take a number of years to thoroughly process and report on. However, as the publications that have emerged from our data set attest to, our approach has facilitated insights into both L2 writing processes and developmental opportunities during tutor-student interaction that would have been difficult to capture without the different ‘lenses’ provided by the QUAN-QUAL methodology. This approach has allowed us to meaningfully examine different levels of the L2 writing complex system.

The epistemological shift seen in recent years, from reductionism and a restricted search for empirical cause-effect relationships towards attempts to understand complex dynamic systems more holistically, has major implications for how we carry out research into L2 writing, as Larsen-Freeman and Cameron (2008: 16) observe: “The theory that we choose to work with, explicitly as researchers and perhaps implicitly as teachers, will dictate how we describe and investigate the world. It controls how we select, out of all that is possible, what to investigate or explain, what types of questions we ask, how data are collected, and what kinds of explanations of the data are considered valid”. Possible implications for future work include:

Investigating Complexity in L2 Writing

1. Collecting data on a wider range of interacting factors in the L2 writing system than is typically done (breadth versus depth).
2. Looking for change or processes leading to changes in writing (longitudinal versus cross-sectional studies).
3. Shifting away from reductionism and at times over-simplified cause-effect explanations of how L2 writing develops.
4. Ecological approaches which account to a larger extent for the *context* in which L2 writing takes place.
5. Deliberate attempts to investigate different levels of the L2 writing system and their interactions.

We hope that our experiences, outlined in this chapter, will encourage others to invest the time, energy, and resources necessary to apply a mixed methods approach to the study of L2 writing processes, as well as other complex dynamic systems. As the writer, Alice Munro (2001) said, “The complexity of things—the things within things—just seems to be endless [...] nothing is easy, nothing is simple.” However, we firmly believe that the rewards of the efforts thus outlined speak for themselves.

References

- Austin, W., Park, C. & Goble, E. (2008). From interdisciplinary to transdisciplinary research: A case study. *Qualitative Health Research*, 18(4), 557-564.
<https://doi.org/10.1177/1049732307308514>
- Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison, Wisconsin: University of Wisconsin Press.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Mahwah, NJ: Erlbaum.
- Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, 6(1), 97-113. <https://doi.org/10.1177/1468794106058877>
- Bryman, A. (2007). Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods Research* 1, 8-22. <https://doi.org/10.1177/2345678906290531>
- Clift, R. (2016). *Conversation Analysis*. Cambridge: Cambridge University Press.
- Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford: Oxford University Press.
- Feilzer, M. Y. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research*, 4, 6-16. <https://doi.org/10.1177/1558689809349691>
- Fogal, G. G. (2019). Tracking microgenetic changes in authorial voice development from a complexity theory perspective. *Applied Linguistics* 40(3), 432-455.
<https://doi.org/10.1093/applin/amx031>
- Ganém-Gutiérrez, A. (2008). Microgenesis, method and object: A study of collaborative activity in a Spanish as a foreign language classroom, *Applied Linguistics* 29(1), 120-148. <https://doi.org/10.1093/applin/amm032>
- Gánem-Gutiérrez, G. A. & Gilmore, A. (2018a). Expert-novice interaction as the basis for L2

Investigating Complexity in L2 Writing

developmental activity: A SCT perspective. *Language and Sociocultural Theory*, 5(1), 21-45.

Gánem-Gutiérrez, G. A. & Gilmore, A. (2018b). Tracking the real-time evolution of a writing event: Second language writers at different proficiency levels. *Language Learning*, 68(2), 469-506. <https://doi.org/10.1111/lang.12280>

Gilmore, A. (2007). *Getting real in the language classroom: Developing Japanese students' communicative competence with authentic materials*. Ph.D. dissertation, Nottingham University, U.K.

Gilmore, A. (2011). "I prefer not text": Developing Japanese learners' communicative competence with authentic materials. *Language Learning*, 61(3), 786-819. <https://doi.org/10.1111/j.1467-9922.2011.00634.x>

Gilmore, A. & Gánem-Gutiérrez, G.A. (forthcoming). Japanese L2 writers' use of Web-based lexicographic tools – Investigating online behavior with digital screen-capture and eye-tracking technologies.

Greene, J.C., Caracelli, V.J. & Graham, W.F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255-274. <https://doi.org/10.3102/01623737011003255>

Hanlon, R.E. & Brown, J.W. (1989). Microgenesis: Historical review and current studies. In A. Ardila, & F. Ostrosky-Solis (Eds.), *Brain organization of language and cognitive processes* (pp. 3-15). Dordrecht, Netherlands: Springer.

Johnson, K. (1995). *Understanding communication in second language classrooms*. Cambridge: Cambridge University Press.

Johnson, R.B., Onwuegbuzie, A.J. & Turner, L.A. (2007). Towards a definition of mixed methods research. *Journal of Mixed Methods*, 1(2), 112-133. <https://doi.org/10.1177/1558689806298224>

Investigating Complexity in L2 Writing

- Jones, R.H. (2008). Good sex and bad karma: discourse and the historical body. In V.K. Bhatia, J. Flowerdew & R.H. Jones (Eds.), *Advances in discourse studies* (pp. 245-57). London: Routledge.
- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. London & New York: Routledge.
- Larsen-Freeman, D. & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford: Oxford University Press.
- Manchón, R. M., Roca de Larios, J. & Murphy, L. (2009). The temporal dimension and problem-solving nature of foreign language composing processes: Implications for theory. In R.M. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research* (pp. 102–129). Clevedon, GBR: Multilingual Matters.
- Miller, K. S., Lindgren, E. & Sullivan, K.P. (2008). The psycholinguistic dimension in second language writing: Opportunities for research and pedagogy using computer keystroke logging. *TESOL Quarterly*, 42, 433-454.
<https://doi.org/10.1002/j.1545-7249.2008.tb00140.x>
- Munro, A. (2001, February 19). Go ask Alice: Alice Munro interviewed by the New Yorker fiction and poetry editor, Alice Quinn. *The New Yorker*, Retrieved from <https://www.newyorker.com/>
- Newman, I., & Benz, C.R. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum*. Carbondale, IL: Southern Illinois University Press.
- Nishino, T. & Atkinson, D. (2015). Second language writing as sociocognitive alignment. *Journal of Second Language Writing*, 27, 37-54.
<https://doi.org/10.1016/j.jslw.2014.11.002>
- Onwuegbuzie, A. J. & Leech, N.L. (2005). On becoming a pragmatic researcher: The

importance of combining quantitative and qualitative research methodologies.

International Journal of Social Research Methodology, 8(5), 375-387.

<https://doi.org/10.1080/13645570500402447>

Scollon, R. (2001). *Mediated discourse: The nexus of practice*. London & New York: Routledge.

Scollon, R. & Scollon, S.W. (2004). *Nexus analysis: Discourse and the emerging internet*. London & New York: Routledge.

Sieber, S.D. (1973). The integration of fieldwork and survey methods. *The American Journal of Sociology*, 78(6), 1335-1359. <https://doi.org/10.1086/225467>

Swain, M. (2010). 'Talking-it-through': Languaging as a source of learning. In R. Batstone (Ed.), *Sociocognitive Perspectives on Language Use and Language Learning* (pp. 112-130). Oxford: Oxford University Press.

Teddlie, C. & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of Mixed Methods in Social and Behavioral Research* (pp. 3-50). Thousand Oaks, CA: Sage Publications.

Trend, M.G. (1978). 'On the reconciliation of qualitative and quantitative analyses: A case study'. *Human Organization*, 37(4), 345-354.

<https://doi.org/10.17730/humo.37.4.181457u6563w1187>

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A. & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In N. Calzolari, K. Choukri, A. Gagemi, B. Maegaard, J. Mariani, J. Odijk, & D. Tapias (Eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation* (pp.1556-1559). Retrieved from www.lrec-conf.org/proceedings/lrec2006/pdf/153_pdf.pdf (Jan 4, 2019).

Investigating Complexity in L2 Writing

Wong Scollon, S. & de Saint George, I. (2012). Mediated discourse analysis. In J.P. Gee & M. Handford (Eds.), *The Routledge handbook of discourse analysis* (pp. 66-78). London: Routledge.

ⁱ Microgenesis refers to ‘the structural development of a cognition (idea, percept, act) through qualitatively different stages [...] from the inception of the cognition to its final representation in consciousness or actualization (expression) in behavior.’ (Hanlon & Brown, 1989, p. 3).

ⁱⁱ Ontogenesis describes the development of an organism within its own lifetime.

ⁱⁱⁱ Phylogenesis describes the development, or evolution, of a particular group of organisms (in this case a specific writing community) over longer time periods.

^{iv} Transcription conventions adapted from Clift (2016, pp. 53–63)

[Indicates a point of overlap onset, whether at the start on an utterance or later
=	Indicates no discernible silence between speaker lines
(0.5)	Silence in tenths of a second (as measured in Elan)
(.)	A micropause
?	Rising intonation
,	Continuing intonation
:	Indicates the prolongation of the sound preceding them
<u>word</u>	Indicates stress or emphasis
Arrows	Indicate sharp rise or fall in pitch

(()) Transcriber’s description of events

° ° Indicates that the talk between the degree signs is softer than the talk around it